Implications of 9 risk prediction models for selecting ever-smokers for CT lung-cancer screening

Hormuzd A. Katki, Ph.D.[1,*,^], **Stephanie A. Kovalchik, Ph.D.**[1,2,^], **Lucia C. Petito, Ph.D.**[1,3], **Li C. Cheung, Ph.D.**[1], **Eric Jacobs, Ph.D.**[4], **Ahmedin Jemal, Ph.D.**[4], **Christine D. Berg, M.D.**[1], and **Anil K. Chaturvedi, Ph.D.**[1,*]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, DHHS, Bethesda, MD, USA

[2]Institute of Sport, Exercise and Active Living, Victoria University, Melbourne, Australia

[3]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

[4]American Cancer Society, Atlanta, GA

Abstract

Background: Lung-cancer screening guidelines now recommend using individualized-risk models to refer ever-smokers for screening. However, different risk-models select different populations to screen. The performance of each risk-model for selecting ever-smokers for screening is unknown.

Objective: To compare the US populations selected for screening by 9 lung-cancer risk-models (Bach/Spitz/LLP/LLPi/Hoggart/PLCO$_{M2012}$/Pittsburgh/LCRAT/LCDRAT), and to examine their predictive performance in 2 cohorts.

Design: Population-based prospective studies.

Setting: United States.

Participants: Model-selected US populations for screening were determined using data from the US-representative 2010–12 National Health Interview Surveys. Model performance was evaluated using data from 337,388 ever-smokers in the NIH-AARP Diet and Health Study (NIH-AARP) and 72,338 ever-smokers in the Cancer Prevention Study II (CPS-II) Nutrition Survey Cohort.

Measurements: Model calibration (number of model-predicted cases divided by number of observed cases (Expected/Observed)) and discrimination (Area-Under-Curve (AUC)).

Results: At a 2.0% 5-year risk-threshold, the models chose US screening populations ranging from 7.6 million to 26 million US ever-smokers. These disagreements occurred because, in both

---

[*]**Corresponding author: HAK:** Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Dr., Room 7E592, Bethesda, MD 20892, Phone: 240-276-7423, Fax: 240-276-7838, katkih@mail.nih.gov.
[^]First two authors contributed equally

validation cohorts, 4 models (Bach/PLCO$_{M2012}$/LCRAT/LCDRAT) were well-calibrated (Expected/Observed from 0.92 to 1.12) and had higher AUCs (0.75 to 0.79) than 5 models that generally overestimated risk (Expected/Observed from 0.83 to 3.69) and had worse AUCs (0.62 to 0.75). The 4 best-performing models also had highest sensitivity at a fixed specificity (and *vice-versa*) and similar discrimination at a fixed risk-threshold. These 4 models better agreed on the population-size selected for screening (7.6–10.9 million) and achieved consensus on 73% of individuals chosen for screening.

Limitations: No consensus on risk-thresholds for screening.

Conclusions: The 9 risk-models chose very different US screening populations. However, 4 risk-models (Bach, PLCO$_{M2012}$, LCRAT, and LCDRAT) best predicted risk and performed best for selecting US ever-smokers for lung-cancer screening.

Keywords

risk modeling; precision prevention; smoking; tobacco

INTRODUCTION

The National Lung Screening Trial (NLST) demonstrated that lung-cancer deaths were reduced 20% by 3 rounds of annual CT screening(1). Annual CT lung-cancer screening is currently recommended by the US Preventive Services Task Force (USPSTF) for smokers aged 55–80 years, with at least 30 pack-years of smoking and at most 15 years since smoking cessation(2). However, recognition is growing that, rather than selecting smokers for screening using simple dichotomized risk-factors, individual-risk calculators that finely account for multiple demographic, clinical, and smoking characteristics could substantially enhance the effectiveness and efficiency of CT-screening programs(3–12),(13). The most recent National Comprehensive Cancer Network guidelines for lung-cancer screening (Version 2.2018) recommend using risk-models to refer ever-smokers for screening(14).

However, many lung-cancer risk-models have been proposed, yet the performance of each risk-model for selecting ever-smokers for screening remains unknown. Independent validation and critical comparison of the statistical properties of multiple risk-models has been limited(15–17). Crucially, there is no evaluation to-date of the comparative performance of risk-models for selecting ever-smokers for screening in the US population. Critical evaluation and comparison is necessary to determine which risk-models could be used clinically.

We applied 9 prominent lung-cancer risk-models to a representative sample of the US population (the 2010–2012 National Health Interview Surveys [NHIS]) to investigate the similarities and differences in the populations of US ever-smokers selected for CT lung-cancer screening by each risk-model. We compared how many US ever-smokers were chosen at fixed risk-thresholds by each risk-model, and the extent of consensus across risk-models on individuals selected for screening, both overall and for subpopulations. We explain our findings by comprehensively evaluating and comparing the statistical performance (calibration and discrimination) of each risk-model in two large, population-

based US cohort studies: the NIH-AARP Diet and Health Study (NIH-AARP) and the American Cancer Society Cancer Prevention Study-II (CPS-II).

METHODS

We describe the risk-model identification process and their characteristics, 2 cohorts used to evaluate the external validity of risk-model predictions, and the application of each risk-model to NHIS data representative of the US population to compare the implications of each risk-model on the US population eligible for screening.

Risk-Model Inclusion

We searched MEDLINE for studies published between January 1, 2000 and December 31, 2016 containing the terms "lung-cancer", "risk", "prediction", and "model". A prediction model was included if (1) it provided a cumulative risk estimate for primary lung-cancer or lung-cancer mortality for at least one time point, (2) was constructed to be valid for general Western populations, (3) did not require biospecimens, (4) did not require CT-screening test results, (5) was internally or externally validated in a disease-free cohort of smokers, and (6) all model parameters were obtainable from the publication or from the authors. When multiple refinements of a model were published, the most recent model was used.

Of the 331 studies identified, 10 risk-models met our inclusion criteria (Supplemental Table S1). Eight models predict lung-cancer risk: Bach[18], Spitz[19], Liverpool Lung Project (LLP) Risk Model [20], Hoggart[21], PLCO Model 2012 ($PLCO_{M2012}$)[12], Liverpool Lung Project Incidence (LLPi) Risk Model [22], Pittsburgh[23], and Lung Cancer Risk Assessment Tool (LCRAT; previously proposed by co-authors[13]). Two models (previously proposed by co-authors) predict lung-cancer mortality: Kovalchik[9] and Lung Cancer Death Risk Assessment Tool (LCDRAT)[13]. We note the Bach lung cancer death model[24], but we could not obtain its parameters. The Spitz and LLP/LLPi models combined case-control data with cancer-registry data; the remaining models were developed in prospective cohorts. The LLP/LLPi and Hoggart models were developed in UK/European samples; all other models were developed in U.S. samples. The Pittsburgh model is a simpler model requiring only four factors. Only Bach, Hoggart, Kovalchik, LCRAT, and LCDRAT account for competing mortality. We previously validated the LCRAT/LCDRAT in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial and the NLST, and the LCDRAT also in the NHIS[13] and a website for them is available (https://analysistools.nci.nih.gov/lungCancerRiskAssessment/). Compared to the Kovalchik model, the LCDRAT model includes more variables and more finely models smoking information. Because the Kovalchik model was not well-calibrated in NIH-AARP and CPS-II (see Supplemental Table S4a-b), we consider it superseded by the LCDRAT. We did not consider the Kovalchik model further, leaving 9 risk-models for consideration.

The **Supplement** details each model. Supplemental Table S2 compares the variables, and their representations, in the models. Most models included additional demographic and smoking-related risk-factors beyond USPSTF-eligibility. The LCRAT, and LCDRAT risk-models also used risk-factors to better account for competing mortality.

Using publicly available risk calculators or software provided by risk-model developers, we confirmed that all models were correctly replicated (Supplemental Table S3). Our publicly available R package *lcmodels* calculates risk from all 9 risk-models (http://dceg.cancer.gov/tools/risk-assessment/lcmodels).

Study Populations for External Model Validation

We used data from the NIH-AAAP and CPS-II cohorts because they were not used to develop any of the prediction models and hence serve as external validation populations. The NIH-AARP cohort (described previously(25)) enrolled 566,398 members of the AARP aged 50–71 years, from six U.S. states and two metropolitan areas with high-quality cancer registries from 1995–1996. After excluding never-smokers and other individuals (see Supplement), 337,388 participants were included in our validation analysis. Incident lung-cancers were identified by linkage to cancer registries. Because age-at-initiation and years quit for former smokers are only known within specified intervals or sometimes unknown, we created multiple-imputation datasets imputing these variables (see Supplement).

The CPS-II Nutrition Survey cohort included 184,194 people enrolled in 1992–1993(26). After excluding never-smokers and other individuals (see Supplement), 72,338 participants were included in our validation analysis. Incident lung-cancers were identified by self-report and verified through released medical records or linkage with state cancer registries. Additional incident lung cancers were detected through linkage with the National Death Index, verified when possible through linkage with state cancer registries.

Nationally Representative Study Population for Lung Cancer Screening Eligibility

The NHIS is a US nationally-representative health survey with annual assessment of the health of the civilian non-institutionalized U.S. population(27). We previously constructed a contemporary U.S. screening population of 43 million ever-smokers ages 50–80 in the NHIS 2010–2012(13). We used multiple-imputation datasets to account for missing information on BMI, race/ethnicity, education, quit years, and cigarettes smoked per day (see Supplement).

In all 3 studies, cause of death was determined by linkage to the National Death Index.

Statistical Analysis for External Validation of the risk-models

Model validity in NIH-AARP and CPS-II was assessed by calibration (both the ratio and the difference of number of model-predicted cases to the number of observed cases (Estimated/Observed and Estimated-Observed)) and discrimination (the receiver operating characteristic (ROC) curve and area-Under-Curve (AUC) statistic). In each cohort, we evaluated calibration overall, across demographic subgroups, USPSTF-eligibility, and across the entire risk range of USPSTF-eligibles (see Supplement). Discrimination was evaluated overall and across USPSTF-eligibility for each model in each cohort. Additionally, we investigated model-performance for three scenarios: 1) fixed specificity (the specificity achieved by the USPSTF criteria in each cohort); 2) fixed sensitivity (the sensitivity achieved by the USPSTF criteria in each cohort); and 3) fixed risk threshold (the threshold that would select the same number of individuals selected by USPSTF criteria in the US population) (see Supplement).

Statistical Analysis for US Population Eligible for Screening

Each model was applied to the 2010–2012 NHIS contemporary cohort to select two types of US populations for screening. First, each model selected a population based on an example 2.0% 5-year lung-cancer risk-threshold or 1.2% 5-year lung-cancer death risk-threshold (both of which select similar numbers of ever-smokers for screening as USPSTF guidelines(13)). We compared the sizes of these 9 populations. Because each model provides a risk pertaining to different years (1,5,6,8.7,10), we distributed 2% risk over 5 years as approximately 0.4% lung-cancer risk per year. Thus, approximately equivalent risk-thresholds are 0.4% 1-year risk, and for example, 2.4% 6-year risk.

Second, each model picked the 8.9 million US ever-smokers at highest model-estimated risk, 8.9 million being the size of the 2010–2012 US-population eligible for screening under USPSTF criteria. We calculated how many individuals were in all 9 model-chosen populations (consensus), or 8 out of 9, and so on.

All analyses were conducted in R, version 3.4.1(28). Analyses of NHIS data used the *survey* package(29).

This study was supported by the Intramural Research Program of the US National Institutes of Health/National Cancer Institute.

RESULTS

Characteristics of study populations

Table 1 compares characteristics of ever-smokers in 4 research cohorts: NIH-AARP, CPS-II, PLCO, NLST, and the nationally representative NHIS. In each of the research cohorts, minorities were under-represented. NIH-AARP and CPS-II had fewer ever-smokers who were currently smoking, 19.5% and 10.3%, respectively, compared to the US population (34.2%), but smokers in these cohorts smoked at greater intensity and had more pack-years of exposure.

Comparative Risk Model Performance in External Populations

Figure 1 and Supplemental Table S4a–b examine the calibration of all models in NIH-AARP and CPS-II. Four models (Bach, PLCO$_{M2012}$, LCRAT, and LCDRAT) were well-calibrated in both cohorts (Expected/Observed from 0.92 to 1.12). The Pittsburgh model was well-calibrated in NIH-AARP (Expected/Observed of 1.01), but underestimated risks in CPS-II (Expected/Observed of 0.83). The other 4 models (Spitz, LLP/LLPi, Hoggart) overestimated risks, with Expected/Observed ranging from 1.18 to 3.69 and Expected-Observed ranging from 55 to 857 per 100,000 per year. For the other 4 models, overestimation was generally smaller for the USPSTF-eligible (Expected/Observed from 0.96 to 2.51) than USPSTF-ineligible (Expected/Observed from 1.64 to 5.76). However, among USPSTF-eligibles, even the well-calibrated models overestimated risk in the highest risk-quintile (except for Bach in NIH-AARP) (Supplemental Figure S1a–t).

For discrimination, the 4 well-calibrated models also had higher AUCs (0.75 to 0.79) than Pittsburgh and the other 4 models (0.62 to 0.75) (Figure 1 and Supplemental Table S4a–b).

Supplemental Figure S2a–b shows the ROC curves for each model in each cohort. The USPSTF criteria achieved a sensitivity/specificity of 0.67/0.67 and 0.48/0.81 in NIH-AARP and CPS-II, respectively. Fixing specificity or sensitivity at the USPSTF criteria level in each cohort, the performance (as measured by sensitivity and specificity, respectively) was highest for LCDRAT but generally similar across the 4 well-calibrated models (Supplemental Tables S5–6). At a fixed risk-threshold of 2.0% risk of lung-cancer incidence (or 1.2% risk of lung-cancer death) over 5 years, the 4 well-calibrated models had similar performance, with no one model achieving both the highest sensitivity and the highest specificity in either NIH-AARP or CPS-II (Supplemental Table S7).

Comparative Risk Model Implications on US Population Eligibility for Lung Cancer Screening

For a screening eligibility risk-threshold of 2.0% lung-cancer risk over 5 years, the 9 risk-models chose populations including 7.6 million to 26 million U.S. ever-smokers (Figure 2). The 4 models that were well-calibrated in NIH-AARP and CPS-II (Bach, PLCO$_{M2012}$, LCRAT, LCDRAT) chose populations ranging from 7.6 million to 10.9 million US ever-smokers. The Pittsburgh model chose a population of 9.6 million. The last 4 models (Spitz, LLP, LLPi, Hoggart) chose populations ranging from 14.5 million to 26 million US ever-smokers. Of note, USPSTF guidelines selected 8.9 million US ever-smokers for screening.

To account for possible miscalibration of risk-models, Supplemental Figure S3a compares populations of 8.9 million US ever-smokers at highest risk selected by each model. Only 20% (1.8M people) of ever-smokers were chosen by all 9 models and USPSTF recommendations. Importantly, 17% of ever-smokers chosen by USPSTF guidelines were chosen by *no* risk-model (1.5M people with an average 5-year risk of lung-cancer incidence by LCRAT of 1.0% and 5-year risk of lung-cancer death by LCDRAT of 0.6% over 5 years). However, among the 4 best-calibrated models (Bach, PLCO$_{M2012}$, LCRAT, LCDRAT), 73% (6.5M people) of ever-smokers chosen by a model were also chosen by the other 3 models (Supplemental Figure S3b).

Performance of risk-models in racial/ethnic and other subgroups

Table 2 details calibration within racial/ethnic subgroups for each risk-model in NIH-AARP for the 4 well-calibrated models overall (Bach, PLCO$_{M2012}$, LCRAT, LCDRAT) and Pittsburgh. Bach and Pittsburgh underestimated risk in African-Americans (Expected/Observed of 0.89 and 0.88 respectively) and overestimated risk in Hispanics (Expected/Observed of 1.34 and 1.29, respectively). PLCO$_{M2012}$ substantially underestimated risk in Hispanic smokers (Expected/Observed: 0.50; 95%CI=0.39–0.65). LCRAT underestimated risk in Asians/others (Expected/Observed: 0.72; 95%CI=0.54–0.97). The models differed in the numbers of US ever-smokers chosen in other subgroups as well: age, current vs. former smoker, family history of lung cancer, or COPD (emphysema/bronchitis) (Supplemental Tables S8a–j and S9)

CONCLUSION

To inform future guidelines for risk-based selection of ever-smokers for CT lung-cancer screening, we compared the US populations selected for screening by 9 risk-models and evaluated model predictions in 2 large US cohorts. The 9 models chose very different sized US populations, ranging from 7.6 million to 26 million ever-smokers, and there was no consensus among all 9 models on which ever-smokers to select for screening. These disagreements were a consequence of the differing predictive performance of the risk-models. Four models (Bach, PLCO$_{M2012}$, LCRAT, LCDRAT) had the best performance, as measured by calibration and discrimination. These 4 models (Bach, PLCO$_{M2012}$, LCRAT, LCDRAT) picked similarly sized US populations (ranging from 7.6 million to 10.9 million ever-smokers) and agreed more on which ever-smokers to select for screening.

Importantly, we found that the four best-performing models (Bach, PLCO$_{M2012}$, LCRAT, LCDRAT) had the highest discrimination overall, highest sensitivity at a fixed specificity and vice-versa, and similar discrimination at a fixed risk-threshold. These observations collectively indicate that any of these models could be used to select smokers in the US population at greatest risk of lung cancer incidence or lung cancer deaths. Each of these models has been validated in external cohorts(13, 16, 17, 30) and have readily-available online risk calculators. Notably, the LCDRAT was previously validated in the NHIS, a representative sample of the US population(13).

This study is the first to compare model-selected populations for CT lung-screening. Model-selected populations differed by nearly 20 million ever-smokers because the largest populations were picked by models that overestimated risk by factors of 2–3 (LLPi, Hoggart, and Spitz). The lack of consensus for selecting ever-smokers at highest risk is due to different risk-factor effects between risk-models. But risk-models are necessary because 17% of those chosen by USPSTF criteria were chosen by *no* risk-model. These findings demonstrate that currently available risk-models have very different implications for whom to offer screening, and underscore the importance of thorough evaluation of risk-model performance in populations being considered for screening. Validation in nationally-representative samples/population is particularly important because research cohorts often recruit volunteers who are healthier than the general population(31). For example, a previous validation study of LCDRAT in the NHIS revealed a hitherto unreported healthy-volunteer effect in the NLST: lung-cancer mortality rates in the NLST were 24% lower than expected(13).

This study represents the most comprehensive external validation of currently available lung-cancer risk-models in the US population, both overall and across key subgroups. Previous efforts have focused on single models (e.g.(30)) or did not validate in cohorts(15), except for a study that compared 4 models (Bach, Spitz, LLP, PLCO$_{M2012}$) in a cohort of Germans(16). Notably, a recent study compared the statistical performance of 3 of the models we compared (Bach, LLP, PLCO$_{M2012}$) in the NLST and PLCO(17), and concluded that models were well-calibrated based on visual examination(17).

Our results highlight the need to improve calibration in important subpopulations even for the 4 overall well-calibrated models. These models overestimated risks in the heaviest smokers and chose different numbers of current smokers. The Bach and Pittsburgh models do not account for race/ethnicity, family history of lung-cancer, or COPD, and chose either too few or too many individuals in comparison to models that account for those variables. PLCO$_{M2012}$ underestimated risks in Hispanics by a factor of 2–3 and chose the fewest numbers of Hispanics. LCRAT and LCDRAT underestimated risk in Asians/others, and chose the fewest number in this subgroup for screening. Thus, even the best-performing lung cancer risk models require refinements to improve prediction in subpopulations.

The models that overestimated risk (LLP, LLPi, Hoggart, and Spitz) either used case-control data or were based on European data. Problems with case-control data could be potential recall-bias of smoking histories in cases or potential lack of population-representativeness of cases and controls. Europe/UK and the US differ in many aspects of tobacco consumption: cigarette composition, anti-tobacco policies, or cultural smoking habits(32). Understandably, European models use no data on US racial/ethnic minorities. Lung-cancer risks may be generally higher in UK/Europe than in the US, even for given smoking histories(33). Conversely, models based on US data may poorly predict risk in Europe/UK (although this was not observed in the German study(16)).

Using only 4 risk factors, the Pittsburgh model outperformed several risk-models, but not the best-calibrated risk-models. Because the Pittsburgh model does not account for quit-years in former smokers, race/ethnicity, family history of lung-cancer, or emphysema/COPD, it may not perform well in those subpopulations. The Pittsburgh model still requires a calculator, and thus should be no easier to use clinically than other risk-models, provided that unknown risk-factors are suitably imputed.

We note our study limitations. The two research cohorts (NIH-AARP and CPS-II) used for validation are not representative of the US population and underrepresent racial/ethnic minorities. Also, both cohorts recruited individuals during the 1990s and thus are not representative of the contemporary US population's smoking exposure. Also, given the lack of consensus for a risk-threshold for the selection of smokers for screening in the US population, we utilized illustrative examples that would select the same number of ever-smokers as the USPSTF criteria.

Very few eligible ever-smokers currently undergo lung-cancer screening, in part because of lack of knowledge about screening, both among providers and the general population, and limited access to screening for those on Medicaid or uninsured.(34) Furthermore, much research remains to be done to develop and evaluate shared decision-making tools and processes for risk-based lung-screening. The University of Michigan lung-screening tool has been in use for over 2 years, using both PLCO$_{M2012}$ and Bach(35). PLCO$_{M2012}$ is downloadable (https://brocku.ca/lung-cancer-risk-calculator/). Another tool, the Lung Cancer Screening Decision Tool(36), uses the Bach lung-cancer death model(24) that we could not validate because we could not obtain its parameters. We have developed an online lung-cancer screening risk-tool, the Risk-based NLST Outcomes Tool (RNOT) (https://analysistools.nci.nih.gov/lungCancerScreening/), that presents not only individual risk of

lung-cancer (using LCRAT), but also individual risk of lung-cancer death (using LCDRAT) and individual risk of false-positive CT screen(13). Our R package *lcrisks* calculates LCRAT/LCDRAT risks (https://dceg.cancer.gov/tools/risk-assessment/lcrisks/) and our R package *lcmodels* calculates risks from all 9 models (https://dceg.cancer.gov/tools/risk-assessment/lcmodels/). A website calculates risks using LCRAT and LCDRAT (https://analysistools.nci.nih.gov/lungCancerRiskAssessment/).

To revise screening guidelines to allow risk-based selection for lung screening, a consensus cost-effective risk-threshold must be determined for screening eligibility. One cited possibility, 1.5% 6-year risk(11), was not based on cost-effectiveness, but by visually noting a drop in prevented lung-cancer deaths in the NLST as a function of risk(9). Our example of 2.0% 5-year risk selects similar numbers of smokers as USPSTF-criteria(13), but this also is not based on cost-effectiveness. Risk-thresholds could be based on lung-cancer incidence or mortality, which chose extremely similar populations for screening.

Ending the epidemic of smoking-related illnesses requires continued progress in smoking cessation and prevention. Effectively and efficiently targeting lung-cancer screening to those at highest risk can further reduce lung-cancer mortality, the leading cause of cancer death. Our findings suggest that 4 lung-cancer risk models are the best models for selecting US ever-smokers for lung-cancer screening. The models should be further refined to improve their performance in subpopulations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS:

References

1. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409. Epub 2011/07/01. doi: 10.1056/NEJMoa1102873. PubMed PMID: . [PubMed: 21714641]

2. Moyer VA, U. S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Annals of internal medicine. 2014;160(5):330–8. doi: 10.7326/M13-2771. PubMed PMID: . [PubMed: 24378917]

3. Bach PB. Raising the bar for the U.S. Preventive Services Task Force. Annals of internal medicine. 2014;160(5):365–6. Epub 2014/01/01. doi: 10.7326/M13-2926. PubMed PMID: . [PubMed: 24379087]

4. Bach PB, Gould MK. When the average applies to no one: personalized decision making about potential benefits of lung cancer screening. Annals of internal medicine. 2012;157(8):571–3. doi: 10.7326/0003-4819-157-8-201210160-00524. PubMed PMID: . [PubMed: 22893040]

5. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and Harms of CT Screening for Lung Cancer: A Systematic ReviewBenefits and Harms of CT Screening for Lung Cancer. Jama. 2012:1–12. . doi: 10.1001/jama.2012.5521. PubMed PMID: .

6. Boiselle PM. Computed tomography screening for lung cancer. JAMA. 2013;309(11):1163–70. doi: 10.1001/jama.2012.216988. PubMed PMID: . [PubMed: 23512063]

7. Gould MK. Clinical practice. Lung-cancer screening with low-dose computed tomography. N Engl J Med. 2014;371(19):1813–20. doi: 10.1056/NEJMcp1404071. PubMed PMID: . [PubMed: 25372089]

8. Gould MK. Who Should Be Screened for Lung Cancer? And Who Gets to Decide? JAMA. 2016;315(21):2279–81. doi: 10.1001/jama.2016.5986. PubMed PMID: . [PubMed: 27179674]

9. Kovalchik SA, Tammemagi M, Berg CD, Caporaso NE, Riley TL, Korch M, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. N Engl J Med. 2013;369(3):245–54. doi: 10.1056/NEJMoa1301851. PubMed PMID: ; PubMed Central PMCID: PMC3783654. [PubMed: 23863051]

10. Tammemagi MC. Application of risk prediction models to lung cancer screening: a review. Journal of thoracic imaging. 2015;30(2):88–100. doi: 10.1097/RTI.0000000000000142. PubMed PMID: . [PubMed: 25692785]

11. Tammemagi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the lung cancer risks at which to screen ever-and never-smokers: screening rules applied to the PLCO and NLST cohorts. PLoS medicine. 2014;11(12):e1001764. doi: 10.1371/journal.pmed. 1001764. PubMed PMID: ; PubMed Central PMCID: PMC4251899. [PubMed: 25460915]

12. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med. 2013;368(8):728–36. doi: 10.1056/ NEJMoa1211776. PubMed PMID: ; PubMed Central PMCID: PMCPMC3929969. [PubMed: 23425165]

13. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and Validation of Risk Models to Select Ever-Smokers for CT Lung Cancer Screening. JAMA. 2016;315(21):2300–11. doi: 10.1001/jama.2016.6255. PubMed PMID: ; PubMed Central PMCID: PMCPMC4899131. [PubMed: 27179989]

14. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Lung Cancer Screening. Version 2.2018. [October 10, 2017]. Available from: https://www.nccn.org/ professionals/physician_gls/pdf/lung_screening.pdf.

15. D'Amelio AM Jr., Cassidy A, Asomaning K, Raji OY, Duffy SW, Field JK, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. British journal of cancer. 2010;103(3):423–9. doi: 10.1038/sj.bjc.6605759. PubMed PMID: ; PubMed Central PMCID: PMCPMC2920015. [PubMed: 20588271]

16. Li K, Husing A, Sookthai D, Bergmann M, Boeing H, Becker N, et al. Selecting High-Risk Individuals for Lung Cancer Screening: A Prospective Evaluation of Existing Risk Models and Eligibility Criteria in the German EPIC Cohort. Cancer Prev Res (Phila). 2015;8(9):777–85. doi: 10.1158/1940-6207.CAPR-14-0424. PubMed PMID: . [PubMed: 26076698]

17. Ten Haaf K, Jeon J, Tammemagi MC, Han SS, Kong CY, Plevritis SK, et al. Risk prediction models for selection of lung cancer screening candidates: A retrospective validation study. PLoS medicine. 2017;14(4):e1002277. doi: 10.1371/journal.pmed.1002277. PubMed PMID: . [PubMed: 28376113]

18. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst. 2003;95(6):470–8. Epub 2003/03/20. PubMed PMID: . [PubMed: 12644540]

19. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst. 2007;99(9):715–26. PubMed PMID: . [PubMed: 17470739]

20. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. British journal of cancer. 2008;98(2):270–6. PubMed PMID: . [PubMed: 18087271]

21. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Ostergaard JN, et al. A risk model for lung cancer incidence. Cancer Prev Res (Phila). 2012;5(6):834–46. doi: 10.1158/1940-6207.CAPR-11-0237. PubMed PMID: ; PubMed Central PMCID: PMC4295118. [PubMed: 22496387]

22. Marcus MW, Chen Y, Raji OY, Duffy SW, Field JK. LLPi: Liverpool Lung Project Risk Prediction Model for Lung Cancer Incidence. Cancer Prev Res (Phila). 2015;8(6):570–5. doi: 10.1158/1940-6207.CAPR-14-0438. PubMed PMID: . [PubMed: 25873368]

23. Wilson DO, Weissfeld J. A simple model for predicting lung cancer occurrence in a lung cancer screening program: The Pittsburgh Predictor. Lung Cancer. 2015;89(1):31–7. doi: 10.1016/j.lungcan.2015.03.021. PubMed PMID: ; PubMed Central PMCID: PMCPMC4457558. [PubMed: 25863905]

24. Bach PB, Elkin EB, Pastorino U, Kattan MW, Mushlin AI, Begg CB, et al. Benchmarking lung cancer mortality rates in current and former smokers. Chest. 2004;126(6):1742–9. doi: 10.1378/chest.126.6.1742. PubMed PMID: . [PubMed: 15596668]

25. Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions : the National Institutes of Health-American Association of Retired Persons Diet and Health Study. Am J Epidemiol. 2001;154(12):1119–25. PubMed PMID: . [PubMed: 11744517]

26. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer. 2002;94(9):2490–501. doi: 10.1002/cncr.101970. PubMed PMID: . [PubMed: 12015775]

27. National Health Interview Survey Hyattsville, MD: National Center for Health Statistics; 2013 [April 23, 2016]. Available from: http://www.cdc.gov/nchs/nhis.

28. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

29. Lumley T Analysis of complex survey samples. J Stat Softw. 2004;9(1):1–19.

30. Cronin KA, Gail MH, Zou Z, Bach PB, Virtamo J, Albanes D. Validation of a model of lung cancer risk prediction among smokers. J Natl Cancer Inst. 2006;98(9):637–40. PubMed PMID: . [PubMed: 16670389]

31. Pinsky PF, Miller A, Kramer BS, Church T, Reding D, Prorok P, et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. Am J Epidemiol. 2007;165(8):874–81. Epub 2007/0½5. doi: kwk075 [pii] 10.1093/aje/kwk075. PubMed PMID: . [PubMed: 17244633]

32. Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980–2012. JAMA. 2014;311(2):183–92. doi: 10.1001/jama.2013.284692. PubMed PMID: . [PubMed: 24399557]

33. Kovalchik SA, De Matteis S, Landi MT, Caporaso NE, Varadhan R, Consonni D, et al. A regression model for risk difference estimation in population-based case-control studies clarifies gender differences in lung cancer risk of smokers and never smokers. BMC Med Res Methodol. 2013;13:143. doi: 10.1186/1471-2288-13-143. PubMed PMID: ; PubMed Central PMCID: PMCPMC3840559. [PubMed: 24252624]

34. Jemal A, Fedewa SA. Lung Cancer Screening With Low-Dose Computed Tomography in the United States-2010 to 2015. JAMA Oncol. 2017. doi: 10.1001/jamaoncol.2016.6416. PubMed PMID: . [PubMed: 28152136]

35. Lau YK, Caverly TJ, Cherng ST, Cao P, West M, Arenberg D, et al. Development and validation of a personalized, web-based decision aid for lung cancer screening using mixed methods: a study protocol. JMIR research protocols. 2014;3(4):e78. doi: 10.2196/resprot.4039. PubMed PMID: ; PubMed Central PMCID: PMC4376198. [PubMed: 25532218]

36. Bach PB. Lung Cancer Screening Decision Tool 2014 [April 23, 2016]. Available from: http://nomograms.mskcc.org/Lung/Screening.aspx.
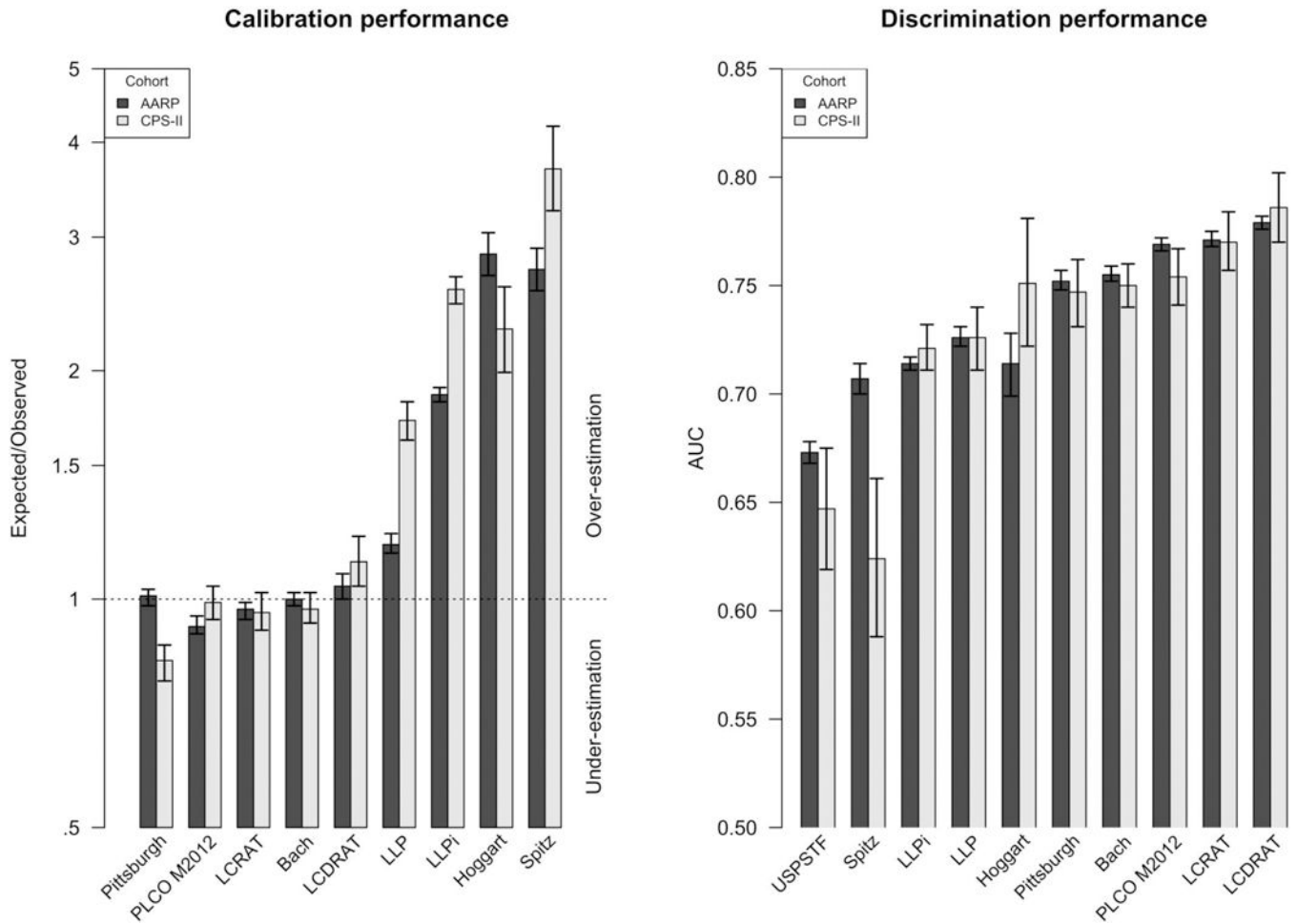
Figure 1: Calibration and discrimination of all 9 models in the NIH-AARP and CPS-II
Abbreviations: NIH-AARP: NIH-AARP Diet and Health Study; CPS-II: Cancer Prevention
Study II Nutrition Survey Cohort; USPSTF: US Preventive Services Task Force. PLCO
M2012: Prostate, Lung, Colorectal, and Ovarian Screening Trial Model 2012; LCDRAT:
Lung Cancer Death Risk Assessment Tool; LCRAT: Lung Cancer Risk Assessment Tool;
LLP: Liverpool Lung Project Risk Model; LLPi: Liverpool Lung Project Incidence Risk
Model.
Footnote: Expected/Observed ratios less than 1 indicate underestimation of risk and
Expected/Observed>1 indicates overestimation. The Area Under Curve (AUC) statistic
examines how well each model discriminates risks between individuals; AUC=1 indicates
perfect model prediction and AUC=0.5 indicates poor model prediction equivalent to purely
random selection.
AUC for the USPSTF criteria represents identifying lung cancer diagnoses within 5 years of
enrolment.

Estimated number of eligible smokers by different models for selecting smokers
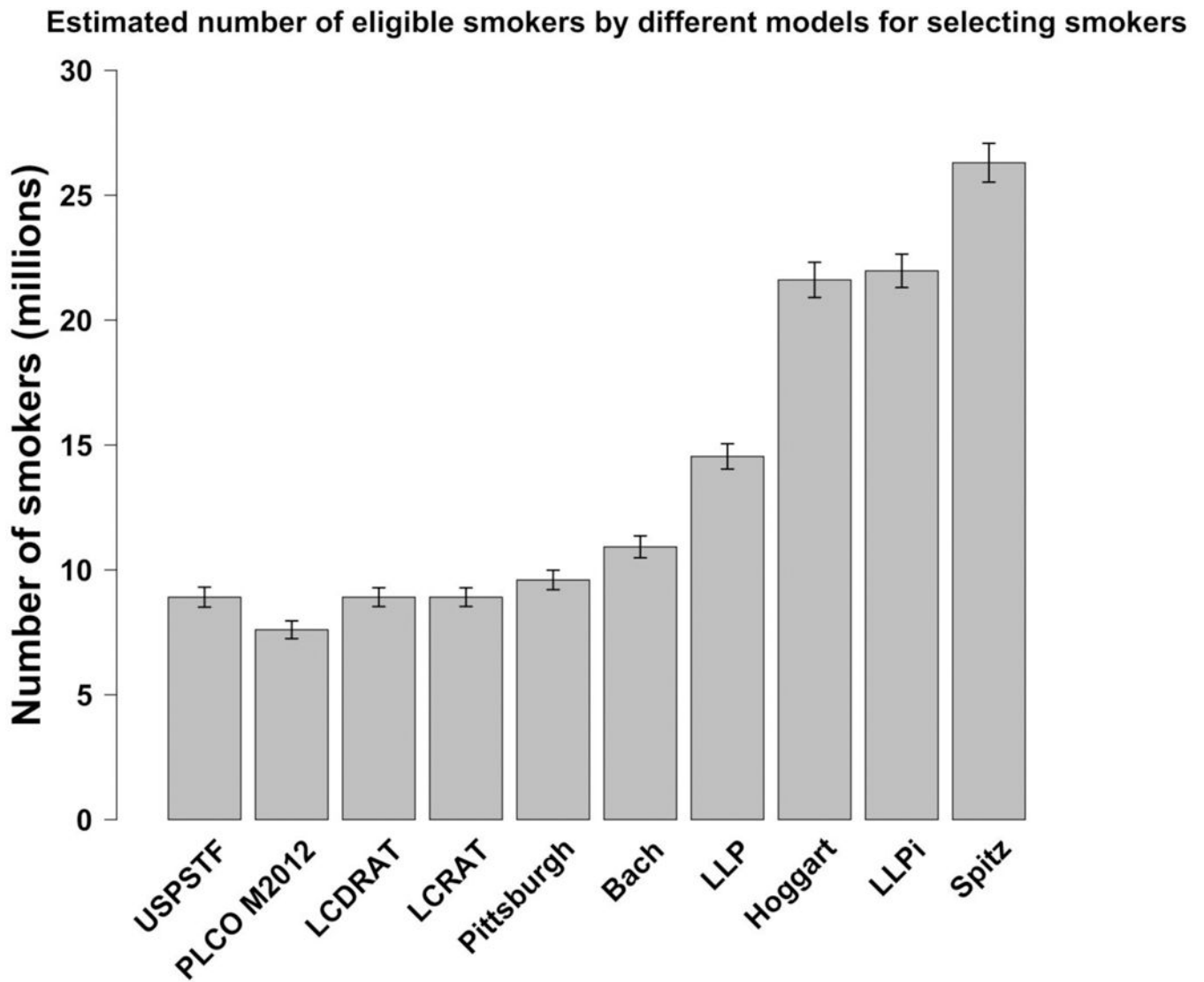
Figure 2: Number of US ever-smokers selected for CT lung-cancer screening by different risk-models at a fixed risk threshold*

Abbreviations: NLST= National Lung Screening Trial; CT= computed tomography; US= United States; USPSTF= United States Preventive Services Task Force; PLCO M2012: Prostate, Lung, Colorectal, and Ovarian Screening Trial Model 2012; LCDRAT: Lung Cancer Death Risk Assessment Tool; LCRAT: Lung Cancer Risk Assessment Tool; LLP: Liverpool Lung Project Risk Model; LLPi: Liverpool Lung Project Incidence Risk Model. Footnote: Our chosen 5-year risk-thresholds for lung-cancer and lung-cancer death (2.0% and 1.2% respectively) yield totals close to the USPSTF-eligible total (8.9 million; dotted line). 2.0% lung-cancer risk over 5-years is equivalent to approximately 0.4% of the population acquiring lung-cancer each year. Approximately equivalent risk-thresholds for those models that do not provide 5-year risks are: 4.0% risk after 10 years for Bach; 0.4% risk after 1 year for Spitz; 3.5% risk after 8.7 years for LLPi; 0.4% risk after 1 year for Hoggart; 2.4% risk after 6 years for PLCOM2012; and 2.4% after 6 years for Pittsburgh.

LCRAT, LCDRAT, and LLP models make projections over 5 years, so no adjustment is needed.

Footnote 2: Risk thresholds needed to select 8.9 million ever-smokers (USPSTF-eligible total) are: 4.8% risk after 10 years for Bach; 1.2% risk after 1 year for Spitz; 3.5% risk after 5 years for LLP; 8.1% risk after 8.7 years for LLPi; 1.3% risk after 1 year for Hoggart; 2.0% after 6 years for PLCOM2012; 2.9% risk after 6 years for Pittsburgh; 2.0% risk after 5 years for LCRAT; and 1.2% death risk after 5 years for LCDRAT.

Table 1.

Baseline Characteristics of ever smokers in NIH-AARP, CPS-II, PLCO, NLST, and NHIS 2010–12

		NIH-AARP	CPS-II	PLCO	NLST	NHIS 2010–12
Sample Size		**337,388**	**72,338**	**79,015**	**53,157**	**18,282**
Number of incident lung cancer		**11,590**	**2,320**	**3,109**	**2,047**	**N/A**
Number of lung cancer deaths		**9,605**	**2,166**	**2,212**	**1,010**	**N/A**
Factor	**Category**	**%**	**%**	**%**	**%**	**%**
Age	<55	12.9	1.1	0.0	0.0	23.9
	55—59	21.8	7.0	34.2	42.7	20.3
	60—64	28.1	23.0	31.1	30.6	19.9
	65—69	33.4	28.9	22.4	17.8	15.1
	70—74	3.8	26.2	12.2	8.8	11.1
	75—79	0.0	11.7	0.0	0.0	8.3
	80	0.0	2.2	0.0	0.0	1.4
Birth year	1900–1918	0.0	3.8	0.0	0.0	0.0
	1919—1929	37.2	48.0	23.3	1.1	0.0
	1930—1934	28.1	27.1	25.6	10.8	6.1
	1935—1939	21.8	16.8	24.9	20.7	9.9
	1940—1948	12.9	4.2	26.2	67.4	27.9
	1949—1962	0.0	0.1	0.0	1.1	56.1
Gender	Male	64.7	55.9	58.2	59.0	55.1
	Female	35.3	44.1	41.8	41.0	44.9
Race/Ethnicity	White, non-Hispanic	93.5	98.1	88.5	90.0	80.3
	African-American	3.6	1.1	5.6	4.4	9.9
	Hispanic	1.6	0.4	2.0	1.8	7.0
	Asian or other	1.3	0.4	3.9	3.9	2.8
BMI (kg/m^2)	Underweight (<18.5)	1.1	1.5	0.7	0.9	1.5
	Normal (18.5–24.9)	32.3	40.6	30.8	27.8	24.4
	Overweight (25.0–29.9)	44.8	41.4	44.3	43.1	21.6
	Obese (30)	21.8	16.5	24.2	28.2	31.4
Prior cancer	No	90.7	82.5	95.4	95.7	84.2
	Yes	9.3	17.5	4.6	4.3	15.8
Family history[a]	0	92.3	95.6	88.8	84.0	91.6
	1	7.5	4.3	10.2	14.8	8.1
	2	0.2	0.0	1.0	1.1	0.3
Smoking status	Current	19.5	10.3	20.2	51.9	34.2
	Former	80.5	89.7	79.8	48.1	65.8
Emphysema	Yes	4.1	7.9	4.3	7.7	6.8
	No	95.9	92.1	95.7	92.3	93.2

		NIH-AARP	CPS-II	PLCO	NLST	NHIS 2010–12
Sample Size		**337,388**	**72,338**	**79,015**	**53,157**	**18,282**
Number of incident lung cancer		**11,590**	**2,320**	**3,109**	**2,047**	**N/A**
Number of lung cancer deaths		**9,605**	**2,166**	**2,212**	**1,010**	**N/A**
Factor	**Category**	**%**	**%**	**%**	**%**	**%**
Chronic Bronchitis[b]	Yes	6.0	3.0	5.9	9.6	7.7
	No	94.0	97.0	94.1	90.4	92.3
Intensity[c]	10	25.3	35.8	25.5	0.0	42.7
	11–20	32.5	38.0	36.6	47.5	39.0
	21–30	19.9	13.6	19.9	27.3	7.6
	31–40	12.6	8.0	11.0	18.1	7.4
	41–60	7.6	4.1	5.7	6.3	2.7
	61	2.1	0.5	1.3	0.8	0.6
Years quit[d]	<1	3.7	0.3	2.1	7.8	3.4
	1—4	8.4	4.8	8.6	27.0	9.3
	5—9	14.2	8.4	11.8	27.2	9.1
	10	73.7	76.5	77.5	38.0	78.1
Age initiated[e]	<15	17.4	7.9	11.4	20.3	18.0
	15—19	53.6	51.7	59.6	50.1	55.8
	20—24	22.2	21.9	21.1	21.5	17.0
	25—29	3.7	4.4	4.0	5.6	4.8
	30—39	2.3	2.6	2.6	2.3	3.0
	40—49	0.6	0.9	0.9	0.2	1.0
	50—59	0.1	0.3	0.3	0.0	0.4
	60—69	0.0	0.0	0.0	0.0	0.0
	70	0.0	0.0	0.0	0.0	0.0
Pack-years[f]	<15	24.7	42.1	25.4	0.0	41.3
	15—29	25.7	23.4	24.7	0.0	24.4
	30—49	24.1	19.7	25.2	52.2	21.6
	50—69	12.8	8.5	12.9	25.3	7.2
	70	12.6	6.4	11.8	22.4	5.5
USPSTF Eligibility[g]	Yes	33.1	19.5	38.5	99.4	21.0
	No	66.9	80.5	61.5	0.6	79.0
Cohort with follow-up (years)	<1	1.7	1.8	1.0	1.2	N/A
	1–4.9	8.3	12.7	6.0	14.4	N/A
	5–5.9	2.2	4.0	1.6	65.4	N/A
	6–8,6	6.5	9.0	6.6	19.0	N/A
	8.7–9.9	4.2	5.2	10.0	0	N/A
	10	77.2	67.2	74.8	0	N/A

Abbreviations: NIH-AARP: NIH-AARP Diet and Health Study; CPS-II: Cancer Prevention Study II Nutrition Survey Cohort; PLCO: Prostate, Lung, Colorectal, and Ovarian Screening Trial; NLST: National Lung Screening Trial; NHIS: National Health Interview Study; USPSTF: US Preventive Services Task Force.

Footnote:

[a.] Number of first degree relatives who had lung cancer.

[b.] Chronic bronchitis: in AARP, imputed from prevalence in control arm of PLCO; in NHIS, question asked as 'told have chronic bronchitis in the past year'.

[c.] Cigarettes per day: in CPS2, NLST, and NHIS, this is an integer-continuous value; in AARP and PLCO, individuals selected a category.

[d.] Years quit: in CPS2, PLCO, NLST, and NHIS, collected as continuous; in AARP, imputed to continuous from categories.

[e.] Age at initiation: in AARP, this is collected only at follow-up questionnaire 2004–2006

[f.] Age 50–80 and smoked at least 30 pack-years.

Missing values (N and %) in the following variables were imputed in AARP: race/ethnicity (3841, 1.1%); bmi (7443, 2.2%); family history of lung cancer (154,420, 45.8%); years quit (131,855, 39.1%); age initiated smoking (169,198, 50.1%).

Missing values (N) in the following variables were reasons for exclusions from the CPS-II cohort: bmi (6,123); cigarettes per day (827); years quit (3,609); age initiated smoking (109); smoking duration (137).

Missing values (N) for age initiated smoking (533) was reason for exclusion from the NHIS 2010–2012 cohort. Family history of lung cancer and smoking intensity among former smokers in NHIS were not collected in the years 2011 and 2012. Missing values (N and %) in the following variables were imputed in NHIS: race/ethnicity (296, 1.6%); bmi (449, 2.5%); family history of lung cancer (13,212, 72.3%); cigarettes per day (8,807 48.2%); years quit (46, 0.3%).

Table 2.

Calibration (Expected/Observed) of the best-calibrated models by race/ethnicity in NIH-AARP.

Model	Race/Ethnicity	Observed Cases /100,000/year	Expected Cases /100,000/year	E/O /100,000/year	E/O Lower 95% CI	E/O Upper 95% CI
LCRAT	White	305.14	296.98	0.97	0.95	1.00
	African-American	308.69	280.88	0.91	0.79	1.05
	Hispanic	176.02	133.60	0.76	0.57	1.01
	Asian/Other	214.15	154.52	0.72	0.54	0.97
LCDRAT	White	175.94	183.26	1.04	1.00	1.08
	African-American	175.47	196.61	1.12	0.93	1.35
	Hispanic	99.01	85.90	0.87	0.59	1.27
	Asian/Other	116.39	92.35	0.79	0.54	1.18
Bach	White	325.15	323.55	1	0.98	1.02
	African-American	296.51	265.37	0.89	0.81	0.99
	Hispanic	174.18	234.23	1.34	1.10	1.65
	Asian/Other	223.46	251.50	1.13	0.92	1.38
Pittsburgh	White	307.03	309.71	1.01	0.98	1.04
	African-American	304.63	267.97	0.88	0.77	1.00
	Hispanic	180.30	232.58	1.29	1.00	1.67
	Asian/Other	201.74	246.08	1.22	0.92	1.61
PLCO$_{M2012}$	White	307.03	284.03	0.93	0.90	0.95
	African-American	304.63	300.43	0.99	0.86	1.13
	Hispanic	180.30	90.16	0.50	0.39	0.65
	Asian/Other	201.74	190.57	0.94	0.72	1.25

Abbreviations: NIH-AARP: NIH-AARP Diet and Health Study; LCRAT: Lung Cancer Risk Assessment Tool; LCDRAT: Lung Cancer Death Risk Assessment Tool; PLCO M2012: Prostate, Lung, Colorectal, and Ovarian Screening Trial Model 2012

Expected/Observed<1 indicates underestimation of risk and Expected/Observed>1 indicates overestimation.

Footnote: Observed and expected cases/100,000/year are the lung cancer incidence per 100,000 individuals per year for LCRAT, Bach, Pittsburgh, and PLCO$_{M2012}$ and lung cancer deaths per 100,000 individuals per year for LCDRAT. Observed rates differ because observed events are lung cancers incidence after 1 year of follow-up for Bach, 5 years of follow-up for LCRAT, and 6 years of follow-up for Pittsburgh and PLCO$_{M2012}$. Observed events are lung cancer deaths after 5 years of follow-up for LCDRAT.